TITLE:

Application of Pair-Wise Discrimination Method to Japanese Consonant Recognition

AUTHOR(S):

Kawahara, Tatsuya; Mizutani, Yoichi; Kitazawa, Shigeyoshi; Doshita, Shuji

CITATION:

Kawahara, Tatsuya ...[et al]. Application of Pair-Wise Discrimination Method to Japanese Consonant Recognition. 音声科学研究 1988, 22: 83-93

ISSUE DATE:

1988

URL:

http://hdl.handle.net/2433/52503

RIGHT:

京都大学学術情報リポジトリ KURENAI 紅
Kyoto University Research Information Repository

STUDIA PHONOLOGICA XXII (1988)

Application of Pair-Wise Discrimination Method to Japanese Consonant Recognition

Tatsuya KAWAHARA, Yoichi MIZUTANI, Shigeyoshi KITAZAWA, Shuji DOSHITA

Abstract

We have proposed pair-wise discrimination method which discriminates multiple classes by combining the results of two-class discriminant analyses performed on the pairs of classes. In this paper, we discuss its application to recognition of all the Japanese consonants. 26 consonant classes for recognition are decided based on the phonetic category and statistical analysis by considering the influence of the following vowels. Effective pairs to discrimination are selected among all the possible ${}_{26}C_2$ pairs. Experimental results show that, as the number of classes increases, the performance of pair-wise discrimination method is outstanding while the conventional one-stage discriminant analysis using common variables for all the classes lowers its recognition rate. It was also confirmed that minimax method is most effective in combining the results of two-class discrimination.

1 Introduction

There exist many approaches to perform automatic speech recognition. Among them phoneme-based recognition is advantageous because the number of phonemes is small and this approach is free from the restriction of the lexical size. In order to realize speaker-independent recognition, we adopt discriminant analysis, or Bayes linear classification, which extracts valid features independent of speakers and environments.

In discriminating multiple classes, the performance of the conventional discriminant analysis lowers due to the use of common variables and a common covariance matrix for all the classes. To solve this problem, we have proposed pair-wise discrimination method. This method utilizes the property that discriminant analysis achieves the highest performance on two-class discrimination. At first, a

Tatsuya KAWAHARA (河原達也) : Master course student, Department of Information Science, Faculty of Engineering, Kyoto University
Yoichi MIZUTANI (水谷陽一) : Master course student, Department of Information Science, Faculty of Engineering, Kyoto University
Shigeyoshi KITAZAWA (北澤茂良) : Associate professor, Department of Computer Science, Faculty of Engineering, Shizuoka University
Shuji DOSHITA (堂下修司) : Professor, Department of Information Science, Faculty of Engineering, Kyoto University

set of pairs of the classes are constructed and discriminant analysis is performed on each pair using the optimal variables and a covariance matrix for the two. Then by combining the results of these analyses, the eventual class to be classified into is decided.

In the previous paper, we reported the experimental results of 9 stop consonant recognition. As the number n of the classes to be discriminated increases, following problems arise.

(1) how to decide classes for recognition which is valid for automatic recognition and is not always same as phonetic classification.

(2) how to select pairs effective to discrimination as the number of pairs increases in the order of n^2.

Considering these problems, we applied pair-wise discrimination method to recognition of all the Japanese consonants. We further investigated how to combine the results of two-class discrimination to obtain the final results.

2 MULTIPLE-CLASS DISCRIMINATION BY DISCRIMINANT ANALYSIS

Input to a classifier is a d-dimensional pattern vector $x=[x_1,x_2, \cdots, x_d]$. Suppose population of class i is normally distributed with mean u_i and covariance $\sum_i$, and suppose further covariance matrices $\sum_i$ $(i=1, \cdots, n)$ are equal to $\sum$. Mahalanobis distance between a given x and mean u_i of class i is

$$D^2_i(x)=(x-u_i)^t \cdot \Sigma^{-1} \cdot (x-u_i)$$

and the probability density function of x is

$$F_i(x)=\frac{1}{(2\pi)^{d/2} \cdot |\Sigma|^{1/2}} \cdot exp\{-D^2_i(x)/2\}$$

A pattern x is classified into the class to which Mahalanobis distance $D_i^2(x)$ is minimum, namely the class whose probability density function $F_i(x)$ is maximum. Since we assume covariance matrices of all the classes to be equal, discriminant function is linear.

In discriminant analysis, all the input variables x_i are not relevant to discrimination. Some variables may be useless or some may be substituted by another. Therefore variables which separate all the classes should be selected. This variable selection is performed statistically. As a result dimension of x is reduced.

In two-class discrimination, optimal variables are selected. As the number of classes increases, however, it is difficult to select variables to maximize Mahalanobis distance among all the classes. Every variable is effective to separate some classes but it may be useless or even harmful to separation of others. In the conventional one-stage discrimination method, variables which contribute to the separation of all the classes on the average are selected. Consequently the variables which can best separate one class from others are not selected if they have no discriminating power for other classes. This causes loss of information of input

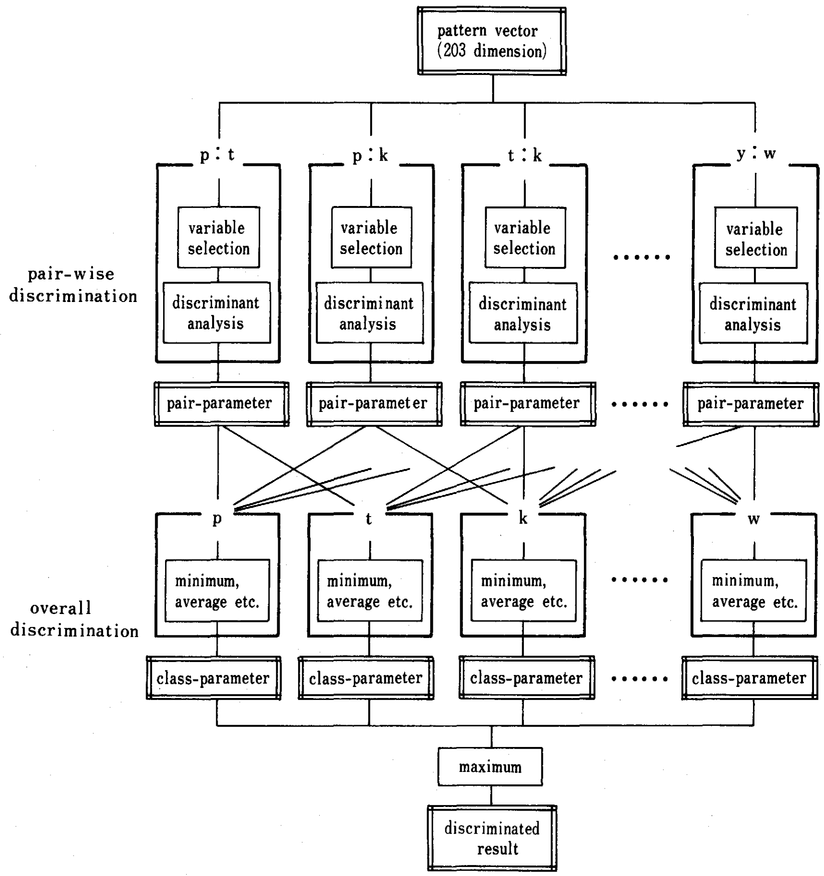

Fig. 1 Flowchart of pair-wise discrimination method.

pattern and lowers the performance of multiple-class discrimination.

3 Pair-Wise Discrimination Method

3.1 Concept of pair-wise discrimination method

To overcome the defect of the conventional one-stage multiple-class discrimination method described in the previous section, we propose pair-wise discrimination method. This method is based on the property that, in two-class discrimination, optimal variables are selected to separate the classes.

At first, a set of pairs of classes are constructed. The number of pairs can be ${}_nC_2$ at most, where n is the number of classes, but it is not necessary to take all the possible pairs. We select such pairs that are effective to discrimination. Details of how to make pairs are discussed in Section 3.4. Next, for each pair of classes, statistical variable selection is performed and two-class linear discriminant analysis is performed using optimal variables which maximize Mahalanobis distance between the two. Here we estimate a specific covariance matrix for each pair which reduces the deviation compared with the analysis using a common covariance matrix for all the classes. By these analyses, pair-parameter is obtained on each pair. Then by combining these pair-parameters, class-parameter is obtained for each classes as the discriminant parameter. A given pattern is classified into the class whose class-parameter is maximum. These steps are illustrated in the Fig. 1

3.2 Pair-parameter of pair-wise discrimination

Since Discriminant analyses are performed on all the pairs using different sets of variables, pair-parameters obtained by them must be normalized to compare each other. In general square distances such as Mahalanobis distance are used as measures of discrimination. Their values are, however, dependent on the dimension of the variable vector and cannot be compared directly. As pair-parameters, here we adopt a posteriori probability, its binarized value (the result of two-class discriminant function) and a upper probability.

(1) a posteriori probability

A posteriori probability is calculated from the probability density function and a priori probability as

$$p(i/x)=\frac{p(i)\cdot F_i(x)}{\sum_{k=1}^{n} p(k)\cdot F_k(x)}$$

For a given pattern x, $p(i/x)$ is the probability that x comes from class i, supposing x comes from any of the class $1\sim$n (in this case $n=2$). Even if a pattern does not belong to any of them, it is forcibly classified into one of them. In pair-wise discrimination, therefore, we cannot conclude that it comes from a class even if its a posteriori probability is high (more than 0.5). We just get negative information that it does not come from the class if its a posteriori probability is low.

(2) binarized value

A binarized value is the result of two-class (1,2) discriminant function. If $p(1/x)$ is higher than $p(2/x)$, the value is 1 for class 1 and 0 for class 2, and vice versa. Just like a posteriori probability, either of two classes gets value 1 even if a pattern does not come from any of them. Calculating this value does not need a covariance matrix nor the probability density function once a linear discriminant function is obtained. So using this value as pair-parameter is advantageous with respect to the amount of computation and storage.

(3) upper probability

A upper probability is a kind of normalized form of the square distance. It is defined by integrating the probability density function outside of the given point and calculated as

$$p'(x|i)=\frac{\Gamma(d/2,\ D_i^2/2)}{\Gamma(d/2)}$$

where d is the dimension of the variable vector and D_i^2 is Mahalanobis distance. We can say that a pattern does belong to a class if its upper probability is high, and it does not if its upper probability is low.

3.3 Multiple-class discrimination method based on pair-wise discrimination

To calculate class-parameter by combining pair-parameters, we adopt following 4 methods: (1) minimax method, (2) average method, (3) majority method and (4) maxmax method. A given pattern is classified into the class whose class-parameter is maximum.

Table 1. The relation between pair-pardmeters and discrimination methods

	a posteriori probability	binarized value	upper probability
minimax method	○	—	○
average method	○	—	○
majority method	—	○	—
maxmax method	—	—	○

○ : the method on the row uses the pair-parameter on the column

(1) minimax method

This method uses a posteriori probabilities or upper probabilities as pair-parameter. It utilizes negative information that a pattern does not come from a class. Class-parameter for each class is defined by the minimum of pair-parameters which are calculated on the pairs containing that class. A pattern is classified into the class least-denied by pair-wise discrimination, namely the class such that probability that a pattern does not belong to it is minimum. This method consequently extracts the crucial pair of classes which are likely to be confused and discrimination is done using best variables to separate them.

(2) average method

This method uses a posteriori probabilities or upper probabilities as pair-parameter. Class-parameter is defined by the average of pair-parameters which are calculated on the pairs containing that class. Class-parameter reflects all the associated pair-parameters.

(3) majority method

This method uses binarized values as pair-parameter. This method is the same as average method, except that instead of probabilities it uses binarized values. It is also same as minimax method if only one class gets 1 as class-parameter, namely if one class is not defeated by the other on any pairs containing it. As explained in the previous section, this method using binarized values is advantageous with respect to the amount of computation and storage.

(4) maxmax method

This method uses upper probabilities as pair-parameter. It utilizes positive information that a pattern does come from a class. Class-parameter is defined by the maximum of pair-parameters which are calculated on the pairs containing that class. A pattern is classified into the class most-supported by pair-wise discrimination.

The relations between pair-parameters and discrimination methods using them are listed in the Table 1.

4 APPLICATION TO SPEAKER-INDEPENDENT CONSONANT RECOGNITION

In this Section, we discuss how to implement pair-wise discrimination on speaker-independent consonant recognition.

4.1 Speech samples and acoustic analysis

Samples examined are all the Japanese consonants followed by one of the five Japanese vowels. The number of these /CV/ syllables is 101. Each syllable was uttered by 17～84 male speakers just once. Speech was recorded at the simple sound-proof booth and digitized to 12 bits at 18.5 kHz sampling rate.

For a given speech sample, reference time-point of consonant part is specified by human observation, and around that point consecutive seven frames are picked out for analysis. For each frame spectrum envelope is calculated by 26th-order LPC analysis and then transformed or merged into 28 variables corresponding to the critical band width. Thus each frame is analyzed to produce 29 variables (28 plus the mean square prediction error of LPC analysis). In total we obtain 203 ($=29*7$ frames) variables, which compose an input pattern vector to discriminant analysis.

4.2 Classification of consonants for recognition

Phonetically there are consonants of */p, t, k, b, d, g, m, n, h, s, z, c, r/* and semivowels of */j, w/* in Japanese. Beside them we add palatalized consonants such as */pj/*. It is not practical to directly use these classification for recognition because the influence of the following vowels makes features of one consonant different. For example, the features of */p/* followed by */a/* are somewhat different from that of */p/* followed by */i/*.

We performed canonical correlation analysis and plot two canonical variables to examine the distribution of pattern vectors of one consonant followed by various vowels. An example is shown in Fig. 2. Viewing this chart, we put */pa, pu, po/* into one class, and */pja, pi, pju, pe, pjo/* into another.

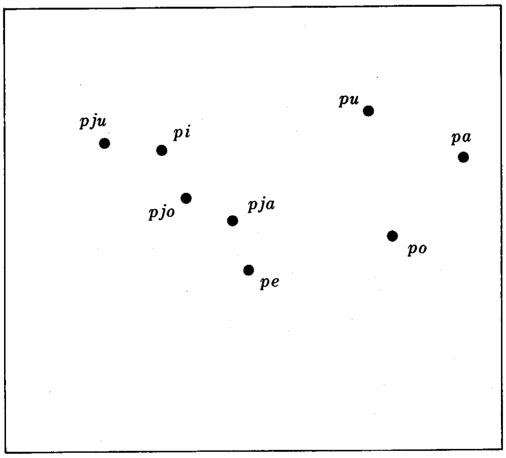

Fig. 2 Plot of pattern vectors by two canonical variables related with consonant /*p*/.

Table 2. Classification of Japanese consonants for recognition

consonant category	syllables containing the consonant	number of samples
ʔ	a, e, o	252
p	pa, pu, po	252
t	ta, ti, tu, te, to	420
k	ka, ku, ko	252
b	ba, bu, bo	252
d	da, di, du, de, do	420
g	ga, gu, go	252
m	ma, mi, mu, me, mo	220
n	na, nu, ne, no	176
ny	nja, ni, nju, njo	176
py	pja, pi, pju, pe, pjo	219
ky	kja, ki, kju, ke, kjo	219
by	bja, bi, bju, be, bjo	219
gy	gja, gi, gju, ge, gjo	219
h	ha, hu, he, ho	68
hy	hja, hi, hju, hjo	68
s	sa, su, se, so	68
sy	sja, si, sju, sjo	68
z	za, zu, ze, zo	68
ts	cu	17
ch	cja, ci, cju, cjo	68
zy	zja, zi, zju, zjo	68
w	wa, u	101
y	ja, i, ju, jo	135
r	ra, ru, ro	51
ry	rja, ri, rju, re, rjo	85

Thus we decided 26 classes for recognition, which is finer classification than phonetic ones listed above. These classes are listed in Table 2. Here ? means the forefront part of the vowels preceded by no phonemes, which is often confused with stop consonants followed by vowels in machine recognition.

Examining this classification, stops followed by /*i*/ and /*e*/ are included in the same category as their palatalized ones, while nasals, fricatives and affricates followed by /*i*/ are the same as their palatalized ones.

4.3 Selecting effective pairs

The possible number of pairs ${}_nC_2$ increases in the order of n^2, which needs enormous computation and storage in multiple-class recognition. It is therefore desirable to select pairs effective to discrimination. We adopt step-wise strategy: At first we construct pairs of classes which seem to have similar features. Then we make recognition experiment using them. Considering its result, we add pairs of classes which are often confused until the recognition rate almost converges.

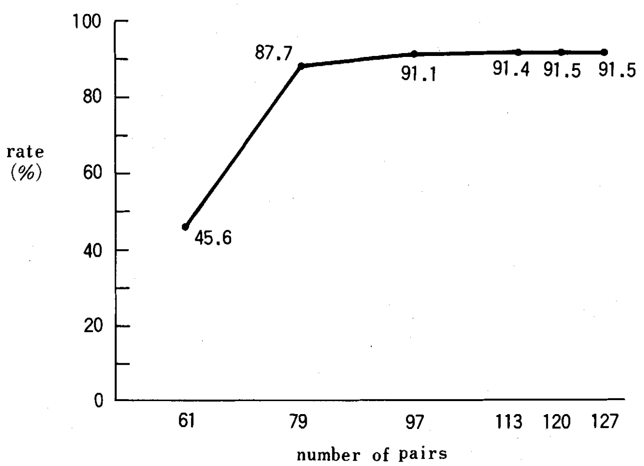

Fig. 3 Relation between the number of pairs and the recognition rate.

The relation between the number of pairs and the recognition rate is shown in Fig. 2. Here we used minimax method using a posteriori probabilities. Thus we selected 120 pairs that is about one third of all the possible 325 ($={}_{26}C_2$). The experimental result shows that there already exist pairs of consonants in 120 which were often confused and there were few confusions between consonants whose pair is not used in discrimination. This tells us that there is much redundancy in all the possible pairs. For example, discrimination between *m-g* is substituted by the pair *m-d* because *g* is similar to *d*. (pair *g-d* is necessary, of course)

5 Experimental Results

5.1 Recognition by one-stage multiple-class discrimination method

We first applied the conventional one-stage discrimination method by discriminant analysis discussed in Section 2. In the experiment, the test data are the same as training data. However we made Jack-knife discrimination which, in recognizing a given sample, uses discriminant function calculated by the samples excluding that one. Jack-knife discrimination realizes open recognition experiment. The recognition rate was 81.0 % and there was much confusion between similar consonants, for example, *p* and *t*. This shows conventional multiple-class discriminanation by discriminant analysis is not effective when the number of the classes to be discriminated is large (26 in this case).

5.2 Recognition by pair-wise discrimination method

Using 120 pairs selected in Section 4.3, we made recognition experiments with 4 methods discussed in Section 3.3. Jack-knife method is performed here, too. The recognition rates are listed in Table 3 and the confusion matrix by minimax method which got highest score is shown in Table 4. In every case, average recognition rate reached almost 90 %. Pair-wise discrimination method reduced recognition errors to half compared with the conventional method.

Table 3. Average recognition rates by each method (percent correct)

	a posteriori probability	binarized value	upper probability
minimax method	91.5	—	89.0
average method	91.5	—	89.3
majority method	—	90.7	—
maxmax method	—	—	89.3

(cf) one-stage multiple-class discrimination method : 81.0%

Minimax method and average method using a posteriori probabilities got much the same score. In minimax method, pair-parameter of one crucial pair, which mostly consists of the discriminated class and the class close to it, affects total discrimination. In average method, too, pair-parameter of such pair significantly affects the average value.

Majority method got a bit worse recognition rate than the others. This is because binarizing pair-parameters causes loss of information for discrimination.

Discrimination methods using upper probabilities could not get good scores. Since an upper probability is an absolute measure, not a relative one, it cannot make clear the difference of the two classes so much and combining it overlooks the results of crucial pairs to distinguish one class from another especially when

Table 4. Confusion matrix by minimax method using a posteriori probabilities

	RATE %	?	p	t	k	b	d	g	m	n	ny	py	ky	by	gy	h	hy	s	sy	z	ts	ch	zy	w	y	r	ry	TOTAL
?	92.5	233	8	0	2	0	0	0	0	0	1	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	252
p	92.1	7	232	7	1	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	252
t	90.0	1	8	378	3	0	4	0	0	1	0	10	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	420
k	93.7	2	3	7	236	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	252
b	90.1	2	16	0	0	227	3	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	252
d	87.9	0	1	27	0	4	369	0	0	1	0	4	0	8	4	0	0	0	0	0	0	0	0	0	0	0	2	420
g	88.1	0	3	1	18	2	1	222	0	1	0	0	0	0	2	0	0	0	0	1	0	0	0	1	0	0	0	252
m	95.0	0	0	0	0	1	0	0	209	8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	220
n	92.0	0	0	0	0	0	0	0	11	162	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	176
ny	98.3	0	0	0	0	0	0	0	2	1	173	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	176
py	92.2	4	0	7	0	0	0	0	0	0	0	202	2	4	0	0	0	0	0	0	0	0	0	0	0	0	0	219
ky	96.3	0	0	6	1	0	0	0	0	0	0	0	211	0	1	0	0	0	0	0	0	0	0	0	0	0	0	219
by	81.7	0	0	0	0	1	4	0	1	0	1	28	0	179	0	1	0	0	0	0	0	0	0	0	0	0	4	219
gy	87.2	0	0	6	0	0	3	0	0	0	1	0	13	2	191	0	0	0	0	0	0	0	1	0	1	0	1	219
h	97.1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	66	0	0	0	0	0	0	0	0	0	0	0	68
hy	100.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	68	0	0	0	0	0	0	0	0	0	0	68
s	95.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	65	3	0	0	0	0	0	0	0	0	68
sy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	63	0	0	0	0	0	0	0	0	0	68
z	92.6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	63	2	1	0	0	0	1	0	68
ts	88.2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	15	0	0	0	0	0	0	17
ch	98.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	67	0	0	0	0	0	68
zy	86.8	0	0	0	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	4	59	0	2	0	0	68
w	98.0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	101
y	97.0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	131	0	1	135
r	90.2	0	0	0	0	0	1	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	1	51
ry	87.1	0	0	0	0	0	2	0	0	2	0	1	2	2	1	0	0	0	0	0	0	0	0	0	0	1	74	85
TOTAL	91.5																											4413

the two classes are much close. Therefore an upper probability is not suitable for pair-wise discrimination method which is a combination of relative discrimination.

Then we review the confusion matrix obtained by minimax method using a posteriori probabilities. We examine the maximum of class-parameter, that is the minimum of a posteriori probabilities of the class to be recognized as. If the value is more than 0.5, the class is supported on all the pairs containing it. We call this sort of discrimination 'assured discrimination'. 7.5 % of 'assured discrimination' are not correct. In these cases, the confusion must have occurred on the first stage, two-class discrimination. On the other hand, 55.8 % of 'non-assured discrimination' are correct, which means that the second stage of pair-wise discrimination which combines the results of two-class discrimination is, to some extent, able to deal with ambiguous results of the first-stage. 97.5 % of all the samples are classified by 'assured discrimination'. This fact shows that, in most cases, two-class discrimination directly affects whole discrimination. Therefore pair-wise discrimination does not lower its recognition performance even if the number of the classes increases. This is confirmed by the fact that the recognition rate of 9 stop consonant recognition is 90.4 %, which is much the same as that of 26 class recognition.

6 Conclusions

We have pointed out the defect of the conventional one-stage multiple-class discrimination by discriminant analysis and proposed pair-wise discrimination method which is based on a set of two-class discrimination. Experimental results of speaker-independent recognition of all the 26 Japanese consonants are encouraging. Pair-wise discrimination method achieved recognition rate of 91.5 % compared with 81.0 % by the conventional method. The advantage of pair-wise discrimination method over the conventional method gets greater as the number of the classes increases when we consider that we could get only 7 % improvement in 9 stop consonant recognition. In combining results of two-class discrimination to get final result, minimax method using a posteriori probabilities was most effective and computationally-cheapest majority method lowered the recognition rate by only 1 %. It was also confirmed that there is no need to construct all the possible pairs of classes. Taking one third of them was enough for 26 consonant recognition.

(Received Oct. 31, 1988)